DOCUMENT RESUME

ED 363 628 TM 020 629

AUTHOR Clonts, Jean G.
TITLE The Concept of Reliability as It Pertains to Data from Qualitative Studies.
PUB DATE Jan 92
NOTE 22p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Houston, TX, January 31-February 2, 1992).
PUB TYPE Information Analyses (070) -- Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Data Collection; Estimation (Mathematics); *Generalizability Theory; Literature Reviews; Models; *Qualitative Research; *Reliability; Research Design; *Research Methodology; Research Problems
IDENTIFIERS *Research Replication

ABSTRACT

This paper presents a review of the literature on reliability in qualitative studies. Reliability is defined as the extent to which studies can be replicated, using the same methods, and getting the same results. It is the degree to which data are independent of the accidental circumstances of the research. The review includes the following three major areas: (1) the use of the qualitative paradigm; (2) the traditional interpretation of reliability; and (3) various strategies for enhancing and insuring reliability. In presenting advantages of a post-paradigmatic view, B. Thompson (1989) notes there are "myriad views of the qualitative paradigm" and urges researchers to be "conscious of the restrictions on insight imposed by their paradigm." Thus, several different perspectives are explored. Strategies are presented to enhance reliability through study design, data collection, and data analysis. Other general categories of strategies that are explored are generalizability theory as an estimate of reliability and the presentation of research as the vehicle of assessing research credibility. Three tables summarize points about research reliability. (Contains 19 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 363 628

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

JEAN G. CLONTS

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Concept of Reliability as It Pertains to Data From Qualitative Studies

Jean G. Clonts
Texas A& M University

BEST COPY AVAILABLE

TM020629

Paper presented at the annual meeting of the Southwest Educational Research Association, Houston, TX, January 30, 1992.

ABSTRACT

This paper presents a review of the literature on reliability in qualitative studies. The review includes three major areas: the use of the qualitative paradigm, the traditional interpretation of reliability, and various strategies for enhancing and insuring reliability. In presenting the advantages of a post-paradigmatic view, Thompson (1989) notes there are "...myriad views of the qualitative paradigm..." (p. 19) and urges researchers to be "...conscious of the restrictions on insight imposed by their paradigm" (p. 4). Thus, several different perspectives are explored within the present review.

As Gronlund (1981, p. 93) notes, "Reliability... provides the consistency that makes validity possible and...indicates how much confidence we can place in our results." It answers the question, "Can independent researchers discover the same phenomena in comparable situations?" (Shimahara, 1988). LeCompte and Goetz (1982) state that while the accuracy of scientific finding involves the issue of validity, reliability involves the replicability of scientific findings.

Although validity and reliability are important components of the objectivity of any research (Kirk & Miller, 1986), reliability is more frequently criticized than validity in qualitative studies (Shimahara, 1988). Reliability and internal validity have a close relationship; they involve the agreement among descriptions of observational phenomena in the same study (LeCompte & Goetz, 1982; Shimahara,1988).

Guba and Lincoln (1981) also note this relationship.

> Since it is impossible to have internal validity without reliability, a demonstration of internal validity amounts to a simultaneous demonstration of reliability. (p.120)

There are differing views of the role and importance of reliability in qualitative studies. LeCompte and Goetz (in press) report that some have questioned whether the reliability of data is a relevant consideration in qualitative studies.

More pragmatically, Kirk and Miller (1972) state "Qualitative researchers can no longer beg the issue of reliability." To

elevate the ethnographic method and the observer to the level of scientific research, the investigator must attend to strategies that maximize validity and reliability (Shimahara, 1988).

To understand reliability, it is necessary to clarify what can be reliable in a qualitative study. Eason (1991) notes that "...reliability is a characteristic of data" (p.84), and Sax (1980, p.261) notes that "...it is...accurate to talk about the reliability of measurements (data, scores and observations)".

However, accepting reliability as a property of measurement information leads to a question of how the presence or absence of that property is determined. Merriam (1988) notes the lack of "...a benchmark by which one can take repeated measures and establish reliability in the traditional sense" (Merriam, 1988, p.170). Also, Goetz and LeCompte (1984) observe that while no study can ever be replicated exactly, because human behavior is not static, reliability directly affects the degree to which study results are credible to others.

The present paper presents a report of a literature review of various methodologists' views of reliability issues in qualitative studies. It examines the qualitative research paradigm, the meaning of reliability in qualitative studies, and presents strategies for increasing reliability, basing the discussion on the seminal article by LeCompte and Goetz (1982) in which they discuss ethnographic research as one variant of the qualitative paradigm.

The Qualitative Paradigm

Qualitative and quantitative research each inform the practice of education, and are considered "...legitimate forms of scientific inquiry" (Borg & Gall, 1989, p. 381). Quantitative research, also called traditional and conventional, originated in the physical and biological sciences (Thompson, 1989). Qualitative research is a newer tradition, and is sometimes called by some naturalistic, subjective, and post-positivistic inquiry (Borg & Gall, 1989; Guba & Lincoln, 1981; Thompson, 1989).

In urging scientists to be "...conscious of the restrictions on insight imposed by their paradigm", Thompson (1989, p.4) quotes Gage (1963):

> Paradigms are models, patterns, or schemata. Paradigms are not the theories; they are rather ways of thinking or patterns for research. (p. 95)

Shimahara (1988) concurs that a paradigm is not a set of rigid rules, but rather a research perspective involving assumptions. A paradigm guides the investigation of issues involving attitudes, values, beliefs, and meaning.

Thompson (1989) notes that while the two paradigms differ in both methodology and purpose, one important difference is in the standards by which truth is tested. One possible component of truth testing is the replication of research findings. In discussing the contribution of qualitative research as unique and distinct from that of quantitative research, LeCompte and Goetz

(1982) state that it answers the question, "What is happening here?".

LeCompte and Goetz (1982) describe the historical relationship of the qualitative paradigm and ethnographic research. Ethnographic research was designed by anthropologists for the study of cultures, and provided the basis for the concepts, values and methods of the qualitative research paradigm. Ethnography is a particular form of qualitative research.

Today, qualitative research is an umbrella term for field study research (Schatzman & Strauss, 1973), and a group of specialized research designs that include case study research (Merriam, 1988), grounded theory (Hutchinson, 1988) and ethnographic studies (LeCompte & Goetz, 1982).

Qualitative research is based on and grounded in the description of observations (Merriam, 1988). Certain characteristics and methods are commonly accepted as appropriate within the qualitative paradigm. These include participant and non-participant observation, a focus on natural settings, the use of particular constructs to structure the research, and avoidance by the investigator of manipulation of the variables within the study (LeCompte & Goetz, 1982). These characteristics define the use of the term qualitative research in the present paper.

Reliability

Reliability is the extent to which studies can be replicated, using the same methods, and getting the same results (LeCompte & Goetz, 1982). It is the degree to which data are independent of

the accidental circumstances of the research, and is dependent upon explicitly described observational proceedings (Kirk & Miller, 1986).

There are other opinions of reliability as the replication of results. Guba and Lincoln (1981), suggesting that it is appropriate to think about "dependability" and "consistency" of results, ask whether others getting the same results would concur that the results make sense. Different results should be regarded as complementary or supplementary and do not refute the earlier study unless there are direct contradictions (Merriam, 1988; Schatzman & Strauss, 1973). On the other hand, replication of grounded theory research, based on the qualitative paradigm, is probably not possible (or even relevant), because the goal of the study is the generation of a new perspective (Hutchinson, 1988).

Eason (1991) discusses reliability as a characteristic of observational and/or measurement data. Comparing the observations of multiple observers of the same phenomenon is recommended to evaluate interobserver reliability (Hutchinson, 1988). However, Rowley (1976) points out what he considers to be the appropriate focus of such investigations:

> What really matters is not the number of times that the particular behavior has been observed, but whether the subjects of the observation have differed consistently in the extent to which they display that behavior. (p. 58)

In discussing case study research, Merriam (1988) reports the views of Scriven (1972): (a) it is possible for a number of persons to experience the same phenomenon, but the observations are not necessarily reliable and (b) increasing the number of observations will not necessarily result in increased reliability.

However, Eason (1991, p.87) suggests a different view of reliability in a quote from Shavelson, Webb, and Rowley (1989, p. 922). "The concept of reliability...is replaced by the broader and more flexible notion of generalizability... Generalizability theory asks how accurately observed scores permit us to generalize about a person's behavior in a defined universe of situations".

Generalizability theory guides estimates in measurement to consider the multiple sources of error that influence scores as well as interaction effects of error (Eason, 1991; Rowley, 1976). Rowley (1976), in proposing a simple method of estimating the reliability of an observational measure by examining the collected data, notes that it is only when an instrument has been used to collect data and the data are manipulated to produce scores "...that we can speak sensibly of reliability" (p. 53).

The concern for reliability of data is clear in Shimahara's (1988) statement that observation, a qualitative method to collect data, can be elevated to scientific research only if the investigator maximizes the validity and reliability of qualitative studies. In 1982, five strategies were proposed by LeCompte and Goetz to enhance reliability in qualitative research: low inference descriptors, multiple researchers, researcher as participant, peer

examination, and mechanically recorded data. The next section of the paper will examine these and additional strategies described in the literature.

Strategies to Enhance Reliability

The strategies presented in the present paper influence one or more of the major phases of qualitative research: the study design, data collection and analysis, and the presentation of findings (LeCompte & Goetz, 1982).

Study design

Triangulation is a critical research design consideration based on the rationale that any single measure of data is fallible as "...a representation of social phenomena" (Fielding & Fielding, 1986, p. 29). The investigator seeks to confirm observations and data-based decisions by examining the data from different sources, either persons or instruments. Fielding and Fielding (1986) describe triangulation as combining methods of data collection (technique triangulation), using more than one researcher data source and acquiring a number of accounts of each event, thereby increasing the researcher's confidence in the accuracy of the data (Merriam, 1988).

Multiple researchers working in the same setting, team observation, and the use of mechanical recording devices (Borg & Gall, 1989; Goetz & LeCompte, 1984, 1982; Merriam, 1988; Shimahara, 1988) provide triangulation of data sources. In other designs, informants provide additional data not readily available to the observer and also to provide other perspective to the researcher.

Peer examination and audit trails are other techniques to insure dependable results. The examination and confirmation of results by peers who independently generated confirming results increase confidence in reliability (Goetz & LeCompte, 1984; Merriam, 1988).

In the case study approach to qualitative research, Merriam (1988) describes the use of independent judges who audit the trail of research: how the data were collected, how categories/constructs were derived and how the decisions were made.

To enhance the value of ethnography as a scientific and legitimate source of knowledge, Goetz and LeCompte (1984) present a system for evaluating the research designs, based on a five dimension scale presented in Table 1. Table 2 lists the categories for evaluation.

INSERT TABLES 1 AND 2 ABOUT HERE.

Data Collection

The selection of data collection methods is a matter of different tools for different jobs (Fielding & Fielding, 1986). The criteria of "informational adequacy" and "efficiency", as proposed by Zelditch (1962, in Fielding & Fielding, 1986), are helpful in the selection process.

Field notes are the primary data collection method in qualitative research and take varied forms, e.g., observations and interviews (both structured and unstructured), questionnaires, photographs, audio and video recordings, survey censuses and

document analysis (Hutchinson, 1988, Kirk & Miller, 1986; LeCompte & Goetz, 1982).

Integral to the general task of data collection is the problem of description during the data collection and in the reporting of results. Goetz and LeCompte (1984) observe that while standardized protocols for data collection are rarely used, apprenticeship and experience in qualitative methods of data collection and analysis are important (Borg & Gall, 1989; LeCompte & Goetz, 1982).

The search for reliability in qualitative observation revolves around the description of the context of the observation (Kirk & Miller, 1986). Noting that the increase in conventions informing the field note format increases reliability, several strategies are suggested by Kirk and Miller (1986). Observations in the setting being studied can be collected by use of instruments of varying structure.

A common method is the use of field notes, which become the base for researcher decisions about the behavior observed and a record that serves during and after the study as a reliable check. The meaningfulness of the notes is enhanced when the questions are recorded.

Kirk and Miller (1986) suggest that field notes must be legible and chronologically ordered. Data should be categorized during the collection time or as soon as possible after data collection (Hutchinson, 1988; Kirk & Miller, 1986). Also, the guide to style suggested in Table 3 clarifies the data entries in field notes (Kirk & Miller, 1986, p. 57).

INSERT TABLE 3 ABOUT HERE.

LeCompte and Goetz (1982, 1984) urge the use of low inference descriptors, noting that shorthand designations should be replaced by careful description. They note that verbatim accounts of behavior and activity and the use of recordings and concrete phrases increase the internal validity (and thus the reliability) of the data.

Several factors of the physical, social and interpersonal context of the setting, as described initially, may change during the collection of data. The process of change must be recorded accurately, because it cannot be reconstructed (LeCompte & Goetz, 1982). An example of change during the data collection occurred in a study by Becker, Geer, Hughes, and Strauss (1961). In a study of the culture of medical students, the investigator noted that the information shared and the student behavior observed when the researcher and the student were alone, changed dramatically when the observations were made of the same medical students as a group.

Mechanically recorded data (photographs, audio and video recordings) are highly accurate observation tools. However, the data are non-codified and, when reviewed by the investigator or others, must be interpreted (LeCompte & Goetz, 1982).

Data analysis

Herriott and Firestone noted "The potential of any study for useful, valid description and generalization depends on the analysts' ability to reduce data to a manageable form without

distortion or loss of meaningful detail" (cited in Thompson, 1989, p. 29). General strategies are needed for analyzing ethnographic data (LeCompte & Goetz, 1982). Merriam (1988) explains qualitative research as a description and explanation of the world as interpreted by those in the world. This implies, by definition, that there will always be multiple interpretations.

LeCompte and Goetz (in press) discuss the recursive nature of analysis within the qualitative paradigm, noting the constant comparison method as proposed by Glaser and Strauss (1967). When the concepts from the analysis are derived from the theoretical framework, the researcher has an "anchor for consistency" (Goetz & LeCompte, 1984, p. 220) that becomes the primary safeguard against unreliability.

Hutchinson (1988), writing about grounded theory as qualitative research, provides a thorough discussion of the circular approach to data analysis. In grounded theory, the constant comparative method of data analysis is the most fundamental methodology. Used to generate theoretical constructs, the process is more definitive when the field notes have been coded or categorized. Hutchinson describes three levels of coding: level 1 notes as small observations, level 2 notes as categorized observation, and level 3 as theoretical constructs.

Regardless of the specific data analytical method used, the cycle of definition and revision requires the researcher to continually examine earlier observations in relation to more recent observations. Long term residence within the research setting and

total immersion in the field enable the recursive nature of qualitative research (LeCompte & Goetz, in press).

Presentation of findings

If reliability of measurements in qualitative research is to be accurately assessed, the investigator must carefully and thoroughly document all procedures (Kirk & Miller, 1986). Schatzman and Strauss (1973) observe that while research is the process of inquiry, the writing of research is the process of communication (p.43), and requires special skills for the thorough communication of research process and findings. A complete description of the research process methods, data collection and analysis enhances reliability (Shimahara, 1988).

Although a strategy to enhance reliability may be employed or considered at a certain phase of the research process, the reader will note one important caveat: the credibility of a study is highly dependent upon the presentation of results. For example, careful and thoughtful decisions made regarding data collecting methods (such as non-participant observation and document analysis) will have little positive influence on the credibility of the study unless each method is thoroughly described in the report of the study. Explaining the assumptions and theory behind the study provides the needed background for evaluating the research purpose, and decisions regarding the investigator's position, the selection of informants and the social context (Butterfield, 1989; Kirk & Miller, 1986; Merriam, 1988).

For example, Neuman (1991) presented a case study of the interaction of learning disabled students with computers and commercial courseware. In the description of methodology, Neuman noted the "...study was conducted and reported according to the principles and procedures of naturalistic inquiry as described" (p. 32).

Additional description about the observer enhances the credibility of a study when field notes are used in data collection (Kirk & Miller, 1986). To place the observation in perspective as a theoretical construct, the reader of the study needs to know the observer, his/her theory of academic commitments, values, behavioral style and experience. When an observation is presented without information about how the observation was collected, it is difficult to place a meaningful interpretation on the observation (Kirk & Miller, 1986).

The social role of the investigator within the research setting determines the flow of information, and therefore influences the type of data and the analysis of the data. The relationship of the researcher (participant or non-participant) with the people being studied must be clearly communicated (Goetz & LeCompte, 1984; Shimahara, 1988). When informants have been used to confirm data or the analysis of data, they must be described carefully and the reasons for their selection explained (Goetz & LeCompte, 1984).

A particular difficulty in communicating research findings is noted by LeCompte and Goetz (1982). In journal-length articles,

the ethnographic researcher is challenged to describe the study fully within the limitations of space.

Summary

This paper presented a review of the literature on reliability in qualitative studies. The review included three major areas: the use of the qualitative paradigm, the traditional interpretation of reliability, and various strategies for enhancing and insuring reliability.

In presenting the advantages of a post-paradigmatic view, Thompson (1989) notes there are "...myriad views of the qualitative paradigm..." (p. 19) and urges researchers to be "...conscious of the restrictions on insight imposed by their paradigm" (p. 4). Indeed, several authors describe a continuum of practice between the qualitative and quantitative paradigms.

LeCompte and Goetz (1989) view ethnography as hypothesis generation and hypothesis verification conducted by experimentation. Hutchinson (1988) speaks of grounded theory research: "Of course, the generalizability of any theory can only be established through verificational studies" (p. 132).

In medical education research, Weinholtz (1989) describes a continuum in which qualitative and quantitative studies each add unique knowledge. The purpose of the initial, qualitative study was to identify effective teaching by attending physicians during teaching rounds in a hospital. One of the questions raised in that study led to a quantitative study to develop and test the

reliability of an instrument for recording effective teaching behaviors of physicians.

In 1982, LeCompte and Goetz analyzed the constructs of validity and reliability in qualitative studies. Questioning the worthiness of the traditional functional definition of reliability as replication of the original research, they suggest "..the generation, refinement, and validation of constructs and postulates may not require replication of (the) situation" (p. 35).

Strategies to enhance reliability have been reviewed, most of which are clarifications of the most appropriate way to utilized commonly used data collection and analysis techniques. However, two other general categories of strategies were mentioned: the use of generalizability theory as an estimate of the reliability of measurement (Eason, 1991; Rowley, 1976) and the presentation itself of research as the vehicle by which research credibility is assessed.

References

Becker, H. Geer, B., Hughes, E. & Strauss, A. (1961). Boys in white: Student culture in medical school. Chicago: University of Chicago Press.

Borg, W.R. & Gall, M.D. (1989). Educational research: An introduction (5th ed.). New York: Longman.

Butterfield, P.S. (1989). Qualitative research strategies and methods in health care settings. Journal of Healthcare Education and Training, 4, 15-21.

Eason, S.J. (1991). Why generalizability theory yields better results than classical test theory: A primer with concrete examples. In B. Thompson (Ed.) Advances in educational research: Substantive findings, methodological developments (Vol. 1, pp. 83-98). Greenwich, CT: JAI Press.

Fielding, N.G. & Fielding J.L. (1986). Linking data. Beverly Hills, CA.: Sage.

Goetz, J.P. & LeCompte. M.D. (1984). Ethnography and qualitative design in educational research. Orlando, FL: Academic Press.

Gronlund, N.E. (1981). Measurement and evaluation in teaching (4th ed.). New York: Macmillan.

Guba, E.G. & Lincoln, Y.S. (1981). Effective evaluation. San Francisco: Jossey-Bass.

Kirk, J. & Miller, M.L. (1986). Reliability and validity in qualitative research. Beverly Hills, CA: Sage.

LeCompte, M.D. & Goetz, J.P. (1982) Problems of reliability and validity in ethnographic research. Review of Educational Research, 52, 31-60.

LeCompte, M.D. & Goetz, J.P. (in press). Qualitative research:What it is, what it isn't. In B.Thompson (Ed.). Advances in social science methodology (Vol. 3). Greenwich, CT: JAI Press.

Merriam, S.B. (1988). Case study research in education. San Francisco, CA: Jossey-Bass.

Neuman, D. (1991). Learning disabled students' interaction with commercial courseware: A naturalistic study. Educational Technology Research and Development, 39, 31-49.

Rowley, G.L. (1976). The reliability of observational measures. American Educational Research Journal, 13, 51-59.

Sax, G. (1980). Principles of educational and psychological measurement and evaluation (2nd ed.). Belmont, CA: Wadsworth.

Schatzman, L. & Strauss, A.L. (1973). Field Research: Strategies for a Natural Sociology. Englewood Cliffs, NJ: Prentice-Hall.

Sherman, R.R. & Webb, R.B. (Eds.) (1988). Qualitative research in education--Focus and methods. New York: Falmer Press.

Thompson, B. (1989). The place of qualitative research in contemporary social science: The importance of post-paradigmatic thought. In B. Thompson (Ed.), Advances in social science methodology (Vol. 1). Greenwich, CT: JAI Press.

Weinholtz, D. (1989). Conducting research using qualitative and quantitative methods. Journal of Healthcare Education and Training, 4, 22-27.

Table 1

Five Dimension Scale for Evaluating Research Designs

APPROPRIATE.....INAPPROPRIATE

CLEAR.....OPAQUE

COMPREHENSIVE.....NARROW

CREDIBLE.....INCREDIBLE

SIGNIFICANT.....INSIGNIFICANT

Note. Adapted from LeCompte and Goetz (1984).

Table 2
Research Design Evaluation Categories

1. Goals of effort and questions asked
2. Conceptual and theoretical framework
3. Overall design or variant that characterizes effort
4. Group providing data
5. Investigator experiences and roles
6. Data collection methods
7. Development of analysis methods
8. Conclusions, interpretations, applications generated

Note. Adapted from LeCompte and Goetz (1984).

Table 3
Guide to Style for Field Notes

" "	verbatim quotes
' '	paraphrase
( )	contextual data and/or research interpretation
< >	angle brackets denoting elements of emic lexicon
_____	solid line, partitions time
/	slash, denoting emic construct

Note. Adapted from Kirk and Miller (1986).